REPORT RESUMES

ED 003 628 56

THE DEVELOPMENT AND EVALUATION OF A PROGRAMMING TECHNIQUE FOR RELATING FRAME DIFFICULTY TO THE ABILITY OF THE LEARNER.
FINAL REPORT, PART I.

BY- MOORE, J. WILLIAM SMITH, WENDELL I.
BUCKNELL UNIV., LEWISBURG, PA.

REPORT NUMBER NDEA-VIIA-1029-1 PUB DATE JUL 65
GRANT OEG-7-48-0070-195
EDRS PRICE MF-$0.18 HC-$3.28 82P.

DESCRIPTORS- PROGRAMED INSTRUCTION, PROGRAMED MATERIALS, *PROGRAMING, *PROGRAMING PROBLEMS, GRADE 8, GRADE 11, *CONCEPT TEACHING, *LEARNING PROCESSES, *ACHIEVEMENT, PROGRAM EVALUATION, LEWISBURG, PENNSYLVANIA

TECHNIQUES OF PROGRAMING WHICH ALLOW FOR A MORE SYSTEMATIC METHOD OF RELATING THE DIFFICULTY OF A FRAME TO THE ABILITY OF THE LEARNER WERE DEVELOPED AND EVALUATED IN FOUR SEPARATE STUDIES. EXPERIMENT ONE TESTED THE HYPOTHESIS THAT ACHIEVEMENT IS A FUNCTION OF THE INTERACTION OF THE ABILITY OF THE LEARNER AND THE AMOUNT OF INFORMATION CONTAINED IN EACH FRAME. AN EXISTING PROGRAM WAS MODIFIED TO CREATE A SMALL-STEP PROGRAM AND A LARGE-STEP PROGRAM. FINDINGS FROM TESTS ADMINISTERED TO 137 EIGHTH-GRADE STUDENTS DID NOT SUPPORT THE HYPOTHESIS. EXPERIMENT TWO SUPPORTED THE SAME HYPOTHESIS, BUT WITH A MORE PRECISE DEFINITION OF FRAME DIFFICULTY. EXPERIMENT THREE WITH 116 EIGHTH-GRADE STUDENTS FOUND THAT FOLLOWING THE ACQUISITION OF ATTRIBUTES, PRACTICE IS OF IMPORTANCE, BOTH DURING THE ASSOCIATION STAGES OF LEARNING AND FOLLOWING CONCEPTUALIZATION, IF THE CONCEPT IS TO BE USEFUL. EXPERIMENT FOUR WITH 120 ELEVENTH-GRADE STUDENTS INDICATED AMONG OTHER RESULTS THAT ACHIEVEMENT IS A FUNCTION OF THE NUMBER OF ASSOCIATIONS PER FRAME WHERE CONSIDERATION IS GIVEN TO THE ABILITY OF THE LEARNER, AND THAT THE LEARNER'S PERCEIVED DIFFICULTY OF A PROGRAM IS A FUNCTION OF HIS LEVEL OF ACHIEVEMENT ON THE PROGRAM. SAMPLES OF MATERIALS USED IN EACH EXPERIMENT WERE PRESENTED AS APPENDIXES IN A SEPARATE DOCUMENT, ED 003 629. (JM)

ED003628

A-1029-1

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

THE DEVELOPMENT AND EVALUATION OF A PROGRAMMING TECHNIQUE FOR RELATING FRAME DIFFICULTY TO THE ABILITY OF THE LEARNER

J. William Moore
Department of Education

Wendell I. Smith
Department of Psychology

Bucknell University
Lewisburg, Pennsylvania

July, 1965

Final Report: USOE, Title VII, Grant No. 7-48-0070-195

TABLE OF CONTENTS

Introduction 1
Review of the Literature 2
Experiment I 8
    Method 8
        Subjects 8
        Materials 9
        Procedure 10
    Results 10
        Error Rates 13
        Program Difficulty 14
    Discussion 16
Experiment II 18
    Introduction 18
    Method 19
        Materials 21
        Procedure 23
        Subjects 25
    Results 26
    Discussion 41
Experiment III 51
    Introduction 51
    Method 52
        Subjects 52
        Materials 52
        Procedure 53
    Results 55
    Discussion 57

TABLE OF CONTENTS (Cont.)

Experiment IV 59
    Introduction 59
    Method 59
        Subjects 59
        Materials 60
        Procedure 61
    Results 62
        Time 63
        Efficiency 66
        Questionnaire 67
    Discussion 68
General Conclusions 73
Bibliography 74
Appendix A 77
Appendix B 78
Appendix C 79
Appendix D 80

The authors of this study wish to acknowledge the significant contributions to the research of Mr. Edward Rosenstock, Susquehanna Valley Program, Bucknell University, and Mr. Roger Davis, Graduate Assistant, Bucknell University.

Appreciation is expressed to the administrators and teachers of the Milton Area Schools, the Sunbury Area Schools, and the Mifflinburg Area Schools for their conscientious participation in the research.

INTRODUCTION

In the past, the development of the majority of programmed materials has involved an intuitive or logical approach to the presentation of information; the characteristics of concepts were identified through inspection and then presented in the number of frames thought necessary for the student to learn the concept. Typically, the program then was tested on students, and appropriate revisions were made. While this procedure is rather effective, it is limited by the fact that the revisions are based on the performance of the students with the lowest achievement in the class. For some students, this approach to programming is unusually redundant, the result often being a reduction in the effectiveness of the program.

The purpose of this investigation was to develop and evaluate techniques of programming which allow for a more systematic method of relating the difficulty of a frame to the ability of the learner.

Four separate studies dealing with related aspects of the problem have been completed. Experiment I tested the hypothesis that achievement is a function of the interaction of the ability of the learner and the amount of information contained in each frame. Experiment II tested the same hypothesis, but with a more precise definition of frame difficulty. Its primary hypothesis was limited to determining the relationship between the number of frames required to learn the attributes of a concept and the level of acquisition of the association stage in concept learning. Experiment III tested the hypothesis that practice of an attribute after its acquisition contributes to the acquisition of the concept.

The last experiment tested the hypothesis that acquisition of a concept is a function of the interaction of the number of attributes associated per frame, the amount of practice of the attributes prior to conceptualization, and the intellectual ability of the learner.

REVIEW OF THE LITERATURE

One of the advantages frequently attributed to programmed instruction is that it provides for individual differences in learners (Klausmeir and Lambert, 1961). Porter (1958) states: "Since each student operates a separate machine, he is free to proceed with the lesson as far and as fast as he can, not held back by less able members of the class." Trow (1963) has stated: "Learning is adapted to individual differences. The simple steps and student pacing reduce individual differences in accomplishment, so that as a consequence, with proper placement and competent programming, students of all abilities can learn." These views reflect Skinner's (1958) belief that the small-step linear form of programming is desirable for all students of all levels of ability. Most authorities do not question the advantages that programmed instruction affords the learner by allowing him to proceed at a rate appropriate to his ability and interest; however, some have suggested that a single, small-step linear program may not be the most desirable (Smith and Moore, 1962; Moore, Smith, and Teevan, 1965; Shay, 1961; Angell and Lumsdaine, 1961). Cline _et al_. (1960) emphasizes this, expressing the view that although the bright student covers the material more quickly, some of the material covered by these students may be needed by only the students of lesser ability. DeCecco (1964), in a discussion of individual differences, states: "A closer inspection of the programs and machines showed them to be rather more innocuous than such extravagant claims indicated. There were self-pacing and feedback, to be sure, but the program sequences were identical for everyone." (p. 346). Although the advantages of self-pacing are obvious, it may be argued that self-pacing provides in only a limited way for individual differences for in the typical linear program all students, without regard for level of ability or previous knowledge of material, are required to complete exactly the same frames in each program. Therefore, it would appear that these advantages that have been ascribed to programmed instruction are true only for a restricted set of conditions and programs.

In an attempt to provide for individual differences, branching or "by-passing" has been used. This approach is based on the assumption that learners can by-pass the material they already know or do not need to review. Campbell (1961) has defined "by-passing" as "...taking extra 'loops' of material at a point in the basic sequence where crucial errors are made or where the learner just bogs down." Campbell (1961) has found that Ss who used a branching program did as well on a criterion test as Ss who followed a fixed sequence, with a saving of time. These findings agree with those of Beane (1962) and of Coulson, Estavan, Melaragno and Silberman (1962). However, others have found no differences in achievement between branching and linear programs, but have found differences in time (Silberman et al., 1961; Glaser and Reynold, 1962; Roe, 1962). These findings would seem to suggest that branching is one technique which may be used to increase the efficiency of programmed materials.

While branching would seem to be a logical approach to providing for individual differences, where efficiency is a primary consideration, it is at best only making limited provisions for the learner who "...may profit from ingesting quite a large amount of material before actively responding to, integrating, or reviewing what has been covered..." (Campbell, 1961). It would appear that branching is at best only a modification of the straight linear form of programming, hence, it makes only limited provisions for individual differences.

The slow progress of programmers in providing more adequately for individual differences may be a function of the techniques used in the preparation of a linear program. Typically, the development of a program begins with the identification of the relevant responses to be taught; frames then are prepared which, it is hoped, will enable the learner to acquire the responses. At this point, the program is tested on a population for which it is deemed appropriate, and revisions are based on the responses of the test population. This procedure is repeated several times until empirical evidence indicates that the learners have acquired the responses to the desired criterion level. A close examination of the revision process reveals that the determination of the effectiveness of the program is made from the performance of students who were lowest achievers in the test population. It would appear that this technique does not provide adequately for the "higher-ability" students to process a greater amount of information; i.e., when compared with the amount processed by the lower-ability students.

One means of compensating for the inadequacies of this procedure has been to manipulate size-of-step. The concept of size-of-step has come to have several connotations during the development of this approach. Homme and Glaser (1959) define size-of-step as the number of steps required to take the learner from the initial to the terminal stimulus-response connections in the program. Carr (1959), while recognizing Homme and Glaser's (1959) definition when size-of-step is used as an independent variable, identifies another meaning of step size: "...when the term is used as a dependent variable, it is usually specified by the percentage of incorrect responses. Thus, if learners make few error responses on a given program, the size-of-step is inferred to be small." Lumsdaine (1959) considers several meanings of the term: "size-of-step may refer to the difficulty of giving a correct response to any item. Also, it may refer to the difference or increment in difficulty between the concept, relationships, or terminology between successive steps or set of steps in a program." (p. 531). In addition the term may refer to the "size of the response segment". (p. 532). "A third sense in which size-of-step sometimes gets used is ... the interval between successive questions to which the learner is asked to respond." (p. 533).

Coulson and Silberman (1960) defined size-of-step in terms of the number of frames in the program and found that small-step forms (104 frames) were significantly superior to a large-step program (56 frames) in amount learned but required significantly more time. As the study was reported, it appears that size-of-step was manipulated through subjective alterations of cues and prompts. Evans, Glaser and Homme (1960) also defined size-of-step in terms of the number of frames in the program. They compared performances of four groups on retention tests and concluded that the groups using the small-step programs (51 and 60 steps) did significantly better on the retention tests than did the other groups (30 and 40 steps). However, they also found that the group using the program with 51 steps did slightly better than the group using the program with 68 steps. In this study, size-of-step was altered through the use of incomplete statements of principles or examples as cues or prompts for the completion of further statements of a rule or example. Smith and Moore (1962) defined step size in terms of the amount of cueing present in a program and found no significant differences in achievement between a large-step and a small-step program in spelling. Size-of-step was manipulated through subjective alternations of the cueing and prompting in the frames.

Campbell (1961) also defined size-of-step in terms of the number of frames in the program and found that students who used a small-step program did significantly better on the criterion measure than students who used a large-step program, regardless of level of ability. However, the short, linear or large-step form differed from the long, linear or small-step form only in the absence of remedial loops. In the long, linear form, S worked straight through the entire program of basic steps and remedial loops with no opportunity to skip any of the remedial loops; i.e., by-passing was not employed. Campbell (1961) also found that students using the small-step program took significantly more time to complete the program than did the students using the large-step program.

It appears that in the majority of the attempts to determine size-of-step, and correspondingly, frame difficulty, the emphasis has been on the relationship of the stimulus portion of a frame to the learner's response. For example, many of the studies suggest that by varying the amount of cueing and prompting in the stimulus portion of a frame, the difficulty of a frame can be varied. However, other investigators have suggested that an adequate definition of the difficulty of a frame must be independent of the student's behavior. Green (1962) states:

> Item analysis cannot give us a direct indication of either the difficulty level or the discriminability of an item that is uncontaminated by other variables. Conventional item analysis is based upon the response of the student to that item and, as such, necessarily incorporates all the other controlling variables that contribute to error or a response. Item analysis, then, does not escape the problem but in some circumstances even compounds it. For this reason, independent description of a program in terms of its density has a considerable advantage over the simple count of the errors that students make to items in that program. Density is independent of the student's behavior. ... It is related to, but not defined by that behavior.

The density of a program has been defined by Green (1962) as the ratio of the total number of different responses in a program to the total number of responses. Green's (1962) concern with developing an adequate description of a program, and correspondingly, of the difficulty of a frame, reflects the inadequacies of existing empirical methods for this purpose.

Program density may be a somewhat more adequate technique for describing the difficulty of a program, independent of the student's behavior, although program density does have limitations. Specifically, program density fails to provide assurances that the responses which are used in determining program density are relevant to the concepts to be learned. Secondly, it does not provide for a measure of the amount of new information which is required to be learned in a given frame.

Thus, it may be argued that an approach is needed which emphasizes the analysis of the content of a frame.

EXPERIMENT I

Objective: To determine the effects of varying the amount of new information presented per frame on the achievement and retention of junior high school students of two levels of ability.

METHOD

Subjects

The Ss were 137 eighth-grade students (63 males and 74 females) enrolled in five mathematics classes.[1] Using the median score from the California Mental Maturity Test, the Ss were classified as "high ability" (above median score) or "low ability" (below the median score). The high ability group was stratified on the basis of sex and randomly assigned to one of three groups. This procedure was repeated for the low ability group. The IQ range for the "high" group was 101 to 129, and for the "low" group, 73 to 100. Equal numbers of Ss from each class were assigned to each group. In order to provide comparable extrinsic motivation for all groups, the Ss were told that a portion of their grade in mathematics would be based on their achievement on the experimental materials.

---

[1] The Ss were eighth-grade students in the Milton Junior High School, Milton, Pennsylvania.

Materials

The first 570 frames of a program in elementary set theory[1] were selected for use in the experiment. In order to investigate the effects of program difficulty on achievement, the program was modified in the following way. The units of information presented per frame in the original program were identified. The number of units per frame was found to vary from one to three. A program was developed which presented the units in the ratio of one unit to one frame. This "small-step" program contained 396 frames. A second program was developed which presented two units per frame. This "large-step" program contained 365 frames. In the original program, the practice items were interspersed through the teaching items and test items. In the two modified programs, the small-step and the large-step, the format was teaching frames, test frames, and practice frames. This presentation was constant for all concepts presented.

The high ability group and the low ability group were randomly assigned to one of the three programs (small-step, original, large-step). Table I presents the experimental design.

TABLE I

Ability	Small-step Program	Original Program	Large-step Program
High			
Low			

---

[1] "An Introduction to Sets, Equations and Inequalities", Reigh, Moore and Smith, New York: McGraw-Hill Book Co.

Procedure

A pretest containing 57 items was administered to all Ss in one session. On the same day, Ss were instructed in the use of the program. The Ss were told to keep the correct answer concealed until they had finished the frame, and they were told that errors on their daily work would not affect their grade. Each classroom was monitored by a teacher who was instructed to provide no information on the content of the program. The error rates on the experimental program were recorded for each S.

When all Ss had completed the experimental materials they were post-tested. On the following day, they completed a questionnaire (D-scale) (See Appendix A) relating to program difficulty. Two months later they were tested for retention. The test of achievement was constructed by the programmer. The pretest, the post-test, and the retention test were identical. In addition, a transfer test was constructed and administered a month after the post-test. To insure the validity of these tests, two items for each concept taught by the program were included. The Kuder-Richardson split-half reliability test resulted in an r of .941, N = 137, for the post-test, and an r of .843, N = 140 for the transfer test. The test is presented in Appendix A.

RESULTS

Achievement

To determine the effects of program difficulty on achievement, analyses of variance were computed on the post-test scores, the retention test scores, and the transfer test scores.

The two-way analyses of variance are reported in Tables II, III and IV.

TABLE II

Two-way analysis of variance of post-test achievement

Source	SS	df	M.S.	F	P
Ability	3642.03	1	3642.03	40.48	< .01
Program	153.19	2	76.60	.85	> .05
Interaction	44.83	2	22.42	.24	> .05
Error	11605.78	129	89.97		
Total	15445.83	134			

TABLE III

Two-way analysis of variance on retention test scores

Source	SS	df	M.S.	F	P
Ability	2600.91	1	2600.91	20.61	< .01
Program	59.04	2	29.52	.23	> .05
Interaction	132.55	2	66.28	.52	> .05
Error	16021.06	127	126.15		
Total	18813.56	132			

TABLE IV

Two-way analysis of variance on transfer test scores

Source	SS	df	M.S.	F	P
Ability	993.07	1	993.07	26.28	< .01
Program	7.36	2	3.68	.10	> .05
Interaction	21.45	2	10.72	.28	> .05
Error	4873.82	129	37.78		
Total	5895.70	134			

The analyses resulted in significant differences at the .01 level between bright and slow Ss in all groups, indicating that Ss above the median in IQ were superior in achievement. Neither of the F's for program difficulty or the interaction of ability and program difficulty were significant. Table V contains the means and standard deviations of post-, retention-, and transfer-test scores.

TABLE V

Post-test Achievement		(Maximum Score = 57 points)	
	Small-Step	Original	Large-Step
High	32.86 N = 22	29.76 N = 21	29.14 N = 21
Low	19.57 N = 28	21.75 N = 20	17.64 N = 25
Retention Achievement		(Maximum Score = 57 points)	
High	21.46	22.15	23.32
Low	13.69	15.95	11.96
Transfer Achievement		(Maximum Score = 31 points)	
High	20.77	20.20	20.14
Low	15.48	14.05	15.00

Error Rates

Error rates were computed for each group and an analysis of variance (Table VI) was done. The groups differed significantly in the number of errors made on the program and in the interaction of ability and programs.

Table VII presents the mean error rates for all groups. It can be observed from this table that the slow Ss had higher error rates on all programs, but their error rates did not vary as a function of program step size. For the brighter Ss, the lowest error rate on the original program can be explained by the difference in programming technique, since the error rates on the small-step and large-step versions were practically the same.

TABLE VI

Two-way analysis of variance on errors made on the program

Source	SS	df	M.S.	F	P
I.Q.	2533.113	1	2533.113	19.03	<.01
Program	.036	1	.036	.0003	>.05
Interaction	10015.177	1	10015.177	74.65	<.01
Between	12548.290	3	4182.763	31.18	<.01
Within	9123.201	68	134.16		
Total	21671.491	71			

TABLE VII

The mean per cent error rates

	Small-Step	Original	Large-Step
High	14.54	8.05	15.90
Low	27.06	28.26	27.01

Program difficulty

Two methods of analysis were used in an attempt to determine the difficulty of the experimental programs. Density: defined as the ratio of the number of different program responses to the total number of program responses. The higher the ratio the greater the number of different responses required, and, possibly, the more difficult the program. Density scores were computed for the large-step and small-step programs using the first sixty frames, the middle sixty frames, and the final sixty frames. Density for the large-step program was .883 and for the small-step program .864. If density is a meaningful measure of difficulty, it would appear that both programs were relatively difficult.

D-scale questionnaire: the D-scale was developed to (a) assess the Ss' perceived difficulty of the programs, and (b) relate this perceived difficulty to step size defined as the amount of information included in a frame. The questionnaire was divided into two sections. Sections I and II differed only in the approach used to determine Ss' perceived difficulty. Section I contained 10 frames selected randomly from the small-step program, and ten comparable frames selected from the large-step program. The Ss were instructed to read each of the 20 frames and rate them on a five-point scale from "very easy" to "very difficult". A difficulty score was computed for each S by assigning a numerical value to his rating of each frame ("very easy" = 1 point, "easy" = 2 points, etc. to "very difficult" = 5 points) and summing these scores over the 20 frames. High scores are indicative of high perceived difficulty.

To determine whether the groups differed in perceived difficulty as a result of ability or previous experience with a program of certain size-of-step, a two-way analysis of variance was computed on the D-scores of Section I. None of the F's reached significance indicating that (a) level of ability did not differentially influence perceived difficulty, (b) previous experience with a program of certain size-of-step did not differentially influence perception of the frames, and (c) the interaction of these two variables did not indicate a difference in the way *S*s viewed the difficulty of the frames.

In order to determine the relationship of perceived difficulty to achievement, *r*'s were computed between post-test scores and D-scores from Section I of the D-scale. Table VIII presents these scores.

TABLE VIII

Correlations between the D-scores and post-test scores for the three E programs for each ability group

	Small-step	Original	Large-step
High	-.643 <.01	-.298 > .10	-.409 > .05
Low	-.428 <.05	-.897 < .01	-.578 <.01

It can be observed from Table VIII that all groups except "high-ability, large-step" and "high-ability, original" exhibited a significant relationship between perceived difficulty and achievement. High achievement tended to be related to perception of the programs as easy.

Section II of the D-scale contained 18 of the frames used in Section I. The frames were presented to Ss in pairs. One frame of the pair had appeared in the small-step program and the other was its counterpart from the large-step program (it contained the same information as the small-step frame and some additional information).

To determine whether Ss could perceive the difference in program step size, they were asked to select the easier frame of each pair.

By use of a chi-square test, it was found that in seven of the nine pairs, a significant number of Ss selected the same frame. In five of these seven pairs, the correct small-step frame was chosen. For the two pairs not reaching significance, the majority of the Ss selected the easier frame.

DISCUSSION

The analysis of the three criterion measures failed to yield significant differences as a function of program step size. These results were consistent with the findings of Goldbeck (1960). Although differences on achievement as a function of ability were obtained, the differences were to be expected inasmuch as high and low ability groups differed by 17 points in IQ. The $\bar{X}$ IQ for the high group was 109 and for the low group 92. The failure to obtain differences on the criterion measures as a function of the number of units of information per frame may be attributable to the failure to perceive the differences between programs. The results of the analysis of Section I of the D-scale and program density would tend to support the hypothesis. Although the results of the analysis of Section II of the D-scale tend to refute the hypothesis, it may be that the Ss differentially perceived frame difficulty only when placed in a forced-choice situation. Since Ss were exposed to only one program during the experiment, it could be argued that Ss' rating of frame difficulty on Section II of the D-scale may not relate directly to the Ss' perception of the difficulty of the experimental program.

An interesting finding of the correlational analyses was the significant correlation between the Ss' perception of the difficulty of the program and their corresponding post-test achievement scores. In four of the six correlations, Ss who perceived the programs as easy tended to score high on the post-test. Only the correlations for the high-ability groups in the original and large-step programs were not significant. Further, the failure to obtain differences on achievement as a function of the type of program may be the result of the difficulty of the programs. An inspection of error rate for the large-step and small-step programs lends some support for this hypothesis.

Although error rate is relatively high for both small-step and large-step programs and for both high and low groups, it can be argued that error rate, in itself, is not a meaningful measure of program difficulty. The results of the analysis of error rates would indicate that the original program was much less difficult for the high IQ group with no corresponding differences in achievement.

Since there is evidence that size-of-step was increased

(396 frames in the large-step program and 365 frames in the small-step program), one could conclude from this study that the size-of-step can be increased without necessarily reducing the effectiveness of the program.

EXPERIMENT II

Introduction

An analysis of Experiment I would suggest, among other possibilities, that the failure to obtain differences as a function of frame difficulty may have been a result of the failure to differentiate between the various stages of learning a concept.

Gagné (1965) has suggested a number of stages in concept learning, one of which he classifies as the association stage. In this stage he suggests that attributes of a concept are learned through verbal associations in much the same way that learning takes place in word association studies. He argues that the association of the attributes of a concept must occur before the conceptualization stage.

Lumsdaine (1960) supports Gagné's emphasis on the association stage of learning a concept with the argument that, "The problem is not to identify the right response or to get it to occur, but to attach it to the right stimulus context." (p. 82)

If Gagné's approach to the learning of a concept can be applied to programmed instruction, it could be argued that the manipulation of step-size may be most meaningful in the association stage of learning the concept. In Experiment I, precise differentiation of frames in terms of stages of learning consistent with Gagné's definition was not adhered to in the attempt to manipulate frame difficulty.

Further analysis of Experiment I would suggest that (a) a more precise definition of increments of information in terms of attributes of concepts is needed, and (b) variation in the number of attributes presented per frame may have been too limited to have a differential effect on achievement.

Objective: The specific purpose of Experiment II was to determine the effects of varying the number of attributes presented per frame on acquisition in the association stage of learning.

METHOD

Experimental Design

A single factor experimental design having repeated measures on the same subject (Winer, 1962) was employed in evaluating the hypotheses, and provided a control on individual differences between the subjects. The experimental design is presented in Table I.

TABLE I

Experimental Design

		E_1	E_2	E_3	E_4	E_5	E_6
High I.Q. (High level of ability)	Subject 1						
	Subject 2						
	⋮						
	Subject n_1						
Low I.Q. (Low level of ability)	Subject 1						
	Subject 2						
	⋮						
	Subject n_1						

The responses of all Ss to the same method of programming might show large variations as a result of individual differences among Ss. In order to provide a control for this source of variation, the treatment effects for each S were measured relative to the average response made by S on all treatments. In this way, each S served as his own control. With this design it was possible to observe each S under each of the treatments.

The experimental conditions were as follows: E_1 presented one attribute of a six-attribute concept per frame; E_2 presented the six attributes of a concept in four frames; E_3 presented the six attributes of a concept in three frames (two attributes per frame); E_4 presented all six attributes in one frame; E_5 presented the six attributes in eight frames; and E_6 used ten frames to present the six attributes.

In order to control for the possible effects of learning set and order, the order of presentation of the concepts under each of the treatments was randomized; i.e., a set consisted of one concept from each of the experimental conditions. Within this set, the order of presentation of the concepts was randomized. This procedure was repeated until the supply of concepts was exhausted. The sets were then randomly assigned, with one restriction: no two concepts from the same classification could appear together. No two subjects received the same order of presentation. Thus, the concepts were randomly assigned among sets, within sets, and the sets were randomly assigned (with the one restriction mentioned earlier).

Three experiments were conducted as tentative tests of the central hypothesis. Experiment A was designed as an initial test of the hypothesis and was designed to give a general indication of the effectiveness of the programming. Experiment B was, for all practical purposes, a replication of Experiment A. The primary purpose of this second experiment was to replicate the study in order to rule out the possibility of systematic differences in treatments as a function of concept difficulty. The first form of the E programs was used in Experiment A, while the second form of the E programs was used in Experiment B. Experiment C attempted to maximize the treatment differences found in the first two studies through the employment of greater variations of the experimental variables. The concepts used in Experiment A were again used as the basis for preparing the programs for Experiment C.

Materials

Six programs were developed for use in the three experiments. The procedure used in developing the programs was as follows: first, a large number of conjunctive concepts were selected from a variety of disciplines, six attributes of each concept were specified for use in describing the respective concepts, and the concepts were then randomly assigned to one of the six experimental conditions. The concepts assigned to each of the experimental conditions were then programmed in the following manner:

In experimental program one (E_1), one attribute was presented per frame. Thus, program E_1 required six information frames to present all six attributes of each concept. A criterion frame, a non-cued frame used to test the level of acquisition of the six attributes, followed the set of six information frames.

The second experimental program (E_2) presented the six attributes in four information frames followed by a criterion frame. In this case, two of the frames presented two attributes per frame, and in two of the frames only one attribute was presented per frame.

Experimental program three (E_3) differed from E_1 only in the fact that two attributes were presented per frame. In this case three information frames and one criterion frame were required.

The number of attributes presented per frame was again increased in the fourth experimental program (E_4). Here, all six attributes were presented in a single frame which was then followed by the criterion frame. Thus, the first four programs differed only in the number of frames used in presenting the attributes.

Program E_5 differed from program E_1 in that two practice frames were included, one after the third and the other after the sixth information frames. The practice frames were included to provide an opportunity to increase the number of times the attributes of the concept were associated prior to the presentation of the criterion frame. In this case, the non-cued frame (practice) required the practice of three attributes which had just been presented in the three frames immediately preceding it. Program E_5 used eight learning frames; i.e., six information and two practice frames and a criterion frame to present the six attributes of each of the concepts included in the program. In this case, opportunity for twelve associations of the attributes of the concept was provided prior to the presentation of the criterion frame.

The preparation of experimental program six (E_6) was exactly the same as program E_5 with one exception. Opportunities for additional associations for each attribute were presented after each two information frames rather than after each three information frames. Further, an additional practice frame providing an opportunity to practice all six attributes was presented at the end of the sequence and prior to the criterion frame. Thus, a total of eighteen associations, three for each attribute, was possible.

The method of programming followed in the preparation of all six programs was to relate the attributes in a meaningful context for S, and then to assume that learning had taken place if S could correctly associate the attributes when they were presented in a different context.

Two forms of programs E_1, E_2, E_3, and E_5 were prepared. Each form contained ten different concepts. Concepts were randomly assigned to each form.

Procedure

All Ss were individually pretested at the beginning of the experiments. The pretests consisted of the criterion item for each concept. No S succeeded in naming all six attributes of any concept in the pretest. Preliminary instructions to S were as follows:

> We are in the process of testing some different types of teaching materials. You will be paid $1.00 an hour for assisting us. First, we need to obtain some information concerning how much you know about the material we are preparing. Please read each question carefully and try to answer it as completely as you can. Take as much time as you wish. If you have any questions, please feel free to ask them.

Following the pretest, Ss were given the following instructions:

> Write your name and age at the top of the sheet. We will now work with the programmed materials. I will hand you a program and time you as you work through it. I want you to mark the time that you start and the time that you finish each program. Read each frame or card and then write your answer on the answer sheet. The answers to each frame are on the back of the card. Cover each answer after you write it and only after you cover it are you to look at the answer on the back of the card. Number each frame as it appears for each program. That is, you will write one, two, three, four, or however many frames you have. You will do this for each program. Be sure that you use the mask to cover your answers. At no time are you to look back at the answer to the previous frames. Here is your first program.

Each concept was handed to S by the investigator. The Ss worked at their own rate and the time required for each frame was recorded.

At the completion of the program, Ss were given a post-test containing the criterion items from the concepts in the experiment. The pretest and the post-test were identical. The Ss were post-tested under the same conditions as existed for the pretest.

To determine the effect of the program variable on Ss' ratings of the difficulty of the concepts, Ss were asked to rate each concept on a difficulty scale constructed by the investigator. The following instructions were given:

> (Name), I want you to indicate on this paper how difficult or easy you think this concept is. You will notice that you can indicate from very hard to very easy.[1]

---

[1] The difficulty rating scale is found in Appendix B.

An attempt to control the extrinsic motivation was made by paying each S one dollar per hour and permitting him to complete as many frames as he wished during any particular experimental period.

Subjects

Experiment A. Fifteen ninth and tenth grade students served as Ss. The ages of the fifteen Ss ranged from 13 to 15 years. Eight Ss had I.Q.'s ranging from 100 to 115 and the mean I.Q. for this group was 108.63, with a standard deviation of 4.24. Seven Ss had I.Q.'s ranging from 115 to 125 with a mean I.Q. of 120.57 and a standard deviation of 3.5. The I.Q.'s were obtained from the Beta form of the Otis Intelligence Test.

Experiment B. The Ss in Experiment B were ten of those from Experiment A; their ages ranged from 13 to 15 years. Five Ss had I.Q.'s ranging from 110 to 115 with a mean I.Q. of 109.20, and a standard deviation of 3.90. The other five had I.Q.'s ranging from 115 to 125; the mean I.Q. was 120.20, with a standard deviation of 2.78.

Procedure for Experiment C

Generally, the procedures followed in Experiment C were the same as those used in Experiments A and B with the following exceptions: The Ss were not run individually in this experiment, hence, the reference to timing in the instructions was deleted. The Ss worked in groups and were pretested and post-tested in these groups.

The programs were not given to S by the investigator as was done in Experiments A and B; rather, Ss were instructed to obtain the programs from a box which contained the programs in a prescribed order. All other conditions were exactly the same.

Subjects

Ninth and tenth grade students served as Ss. The ages of the fifteen Ss ranged from 13 to 15 years. Seven of the Ss had I.Q.'s ranging from 100 to 115 with a mean I.Q. of 103.86 and a standard deviation of 5.49; eight of the Ss had I.Q.'s ranging from 115 to 125 with a mean I.Q. of 121.25 and a standard deviation of 4.56.

RESULTS

An analysis of covariance was performed for each of the three experiments to test the following two hypotheses: (1) achievement, as measured by the responses to the criterion items, is a function of (a) the number of associations of attributes per frame, (b) the level of student ability, or (c) the interaction of (a) and (b); (2) achievement, as measured by the responses to the criterion items, is a function of (a) the number of associations of attributes per concept, (b) the level of student ability, or (c) the interaction of (a) and (b). The responses to the individual criterion items were summed over the concepts under each experimental condition for each S. The analyses of covariance employed the scores from the pretest, obtained in the same manner as the post-test scores, as a control variable.

The results of the analyses for each of the experiments are presented in Table II.

TABLE II

Analyses of Covariance on the Total Number of Correct Responses to the Criterion Items

Experiment A

Source	SS_x	SS_{xy}	SS_y	SS'_y	df	MS'_y	F
I.Q.	26.7853	69.9996	182.9342	1017.9712	1	1017.9712	NS
Residual	1139.6114	10087.7163	108301.0000	19005.5441	12	1583.7953	
Treatment	173.7680	322.8989	701.5099	231.4070	3	77.1357	14.3266**
IQxTreatment	10.8357	20.3662	44.9742	21.6338	3	7.2113	1.3394^{NS}
Residual	126.9643	54.0179	206.8929	204.5947	38	5.3841	

Experiment B

Source	SS_x	SS_{xy}	SS_y	SS'_y	df	MS'_y	F
I.Q.	.0250	1.75	122.5000	130.7243	1	130.7243	2.8119^{NS}
Residual	25.5000	59.8500	465.9000	325.4285	7	46.4898	
Treatment	67.4750	89.6000	245.8000	203.5792	3	67.8597	7.3597**
IQxTreatment	.8750	3.7500	72.9000	71.8112	3	23.9371	2.5961^{NS}
Residual	78.9000	9.8500	213.3000	212.9703	23	9.2205	

Experiment C

Source	SS_x	SS_{xy}	SS_y	SS'_y	df	MS'_y	F
I.Q.	.2927	1.0815	223.5635	282.1606	1	282.1606	NS
Residual	1358.8218	15041.5202	187992.5632	21490.4553	12	1790.8713	
Treatment	161.7795	117.4915	196.5004	111.4815	4	27.8704	6.1285**
IQxTreatment	15.0529	19.0378	39.1666	18.1022	4	4.5256	NS
Residual	128.6785	101.8964	312.6214	231.9329	51	4.5477	

**$p < .01$

For all experiments, the treatment effects were significant. (For Experiment A, F = 14.3266, df = 3/38; Experiment B, F = 7.3597, df = 3/23; Experiment C, F = 6.1285, df = 4/51.) The effect of I.Q. and the effect of the interaction of I.Q. and treatments were not significant in any of the experiments. To determine which means differed as a function of the treatment effects, the means for each of the experimental conditions were compared using Scheffe's Method for Multiple Comparisons. In Scheffe's Method, the critical value for each comparison differs for a given level of significance; therefore, only the levels of significance are indicated. The results of the comparisons for each of the three experiments appear in Table III.

TABLE III

Comparison of Mean Scores - Responses to the Criterion Items

Treatment Means					
Experiment A	$\bar{X}_{E_i}$	$\bar{X}_{E_i}-\bar{X}_{E_3}$	$\bar{X}_{E_i}-\bar{X}_{E_2}$	$\bar{X}_{E_i}-\bar{X}_{E_1}$	
$\bar{X}_{E_5}$	49.3320			5.2511**	
$\bar{X}_{E_1}$	44.0809	2.3916	-1.3504		
$\bar{X}_{E_2}$	45.4313	3.7420**			
$\bar{X}_{E_3}$	41.6893				
Experiment B					
$\bar{X}_{E_5}$	52.5527			7.2371**	
$\bar{X}_{E_1}$	45.3156	-2.1003	-2.6002		
$\bar{X}_{E_2}$	47.9158	.4999			
$\bar{X}_{E_3}$	47.4159				
Experiment C	$\bar{X}_{E_i}$	$\bar{X}_{E_i}-\bar{X}_{E_4}$	$\bar{X}_{E_i}-\bar{X}_{E_3}$	$\bar{X}_{E_i}-\bar{X}_{E_2}$	$\bar{X}_{E_i}-\bar{X}_{E_1}$
$\bar{X}_{E_6}$	55.0385				1.5665
$\bar{X}_{E_1}$	53.4720	-1.3534	.7661	1.6612	
$\bar{X}_{E_2}$	51.8108	-3.0146*	-.8951		
$\bar{X}_{E_3}$	52.7059	-2.1195			
$\bar{X}_{E_4}$	54.8254				

$*p < .05$, $**p < .01$

To determine whether the variations in the number of associations per frame produced a statistically significant variation in the achievement of the learner, the following means were compared. In Experiments A and B, the means for programs E_1, E_2, and E_3 were compared, and in Experiment C the means for programs E_1, E_2, E_3, and E_4 were compared. Comparisons were limited to these experimental conditions because programs E_5 and E_6 introduced practice frames and would thus, if compared, for example, to program E_3, yield uninterpretable results.

To determine whether the variation in the number of associations of attributes per concept produced a statistically significant variation in the achievement of the learner, the following means were compared. In Experiments A and B, the comparisons were made between the means for programs E_1 and E_5 and in Experiment C, the comparisons were made between the means for programs E_1 and E_6. Table III presents the results of the comparisons for both hypotheses.

Comparisons of the means (Table III) reveal that in Experiment A, program E_2 resulted in significantly greater achievement than program E_1.

In Experiment B, other comparisons of the means did not result in significant differences (E_1, E_2, and E_3 programs), although the same trend in number of associations of attributes as was observed in Experiment A persists in Experiment B.

In Experiment C, there were significant differences as a function of variations in the number of associations of attributes per frame. The only significant difference in this case is between programs E_2 and E_4, with a higher mean for program E_4 than for program E_2.

There were significant differences as a result of variations in the number of associations of attributes per concept in two of the three experiments. In Experiment A, program E_5 resulted in significantly greater achievement than program E_1. The same difference is observed in Experiment B. A comparison of the means of programs E_6 and E_1 in Experiment C resulted in no significant differences between them.

It will be noted that the only comparisons made among means in evaluation of this last hypothesis were among means of the E_1, E_5 and E_6 programs. Any other comparisons would be uninterpretable because of the nature of the programming in the other treatments. Specifically, programs E_5 and E_6 differed from program E_1 only in the number of associations of attributes per concept, while they differed in two ways from programs E_2, E_3 and E_4, (1) in the number of attributes per concept and (2) in the number of attributes per frame.

An analysis of covariance was completed to test the following hypotheses of each of the respective experiments: (1) achievement, as measured by the responses to the post-test, is a function of (a) the number of associations of attributes per frame, (b) the level of student ability, or (c) the interaction of (a) and (b); (2) achievement, as measured by the responses to the post-test, is a function of (a) the number of associations of attributes per concept, (b) the level of student ability, or (c) the interaction of (a) and (b). The analyses of covariance employed the Ss' scores on the pre-test as the control variable.

The results of the analyses for each of the experiments are presented in Table IV. In all experiments, the treatment effects were significant (for Experiment A, F = 6.1386, df = 3/38; Experiment B, F = 9.1641, df = 3/23; Experiment C, F = 17.3788, df = 4/51). Only in Experiment B was the effect of I.Q. significant (F = 5.6585, df = 1/7). Other effects were not significant. The means for each of the experimental conditions were compared using Scheffe's Method for Multiple Comparisons; the same comparisons that were made for the criterion items were performed with the means from the post-test. The results appear in Table V.

In Experiment A, there were significant differences as a result of the variation in the number of associations of attributes per frame. The means of programs E_1 and E_2 were significantly *greater than the mean for program* E_3. *Note that there* was not a significant difference between the means for programs E_1 and E_2.

In Experiment B, there were significant differences between the means of programs E_1 and E_2 with program E_2 producing the larger mean, as well as between the means of programs E_2 and E_3 with program E_2 again having the larger mean. There were no significant differences between the means of programs E_1 and E_3.

In Experiment C, the means of programs E_1, E_2, and E_3 were significantly greater than the mean for program E_4. The only other significant comparison was between the means for programs

TABLE IV

Analyses of Covariance of the Total Number of Correct Responses to the Post-test for Experiments A, B, and C

Source	SS_x	SS_{xy}	SS_y	SS'_y	df	MS'	F
Experiment A							
I.Q.	26.7853	151.6950	859.1066	23.3733	1	23.3733	NS
Residual	1139.6114	5377.1320	29715.2347	4343.7751	12	361.9813	
Treatment	173.7680	374.5147	821.2183	167.5264	3	55.8421	6.1386**
IQxTreatment	10.8357	3.0830	48.2229	49.1434	3	16.3811	1.8007^{NS}
Residual	126.9643	90.0000	409.4778	345.6803	38	9.0969	
Experiment B							
I.Q.	.0250	3.1500	396.9000	401.5127	1	401.5127	5.6585*
Residual	25.5000	20.2000	512.7000	496.6984	7	70.9569	
Treatment	67.4750	109.9000	616.4000	483.1689	3	161.0563	9.1641**
IQxTreatment	.8750	2.8500	29.1000	26.3625	3	8.7875	NS
Residual	78.9000	40.0000	424.5000	404.2212	23	17.5748	
Experiment C							
I.Q.	.2927	2.1250	15.4822	6.9089	1	6.9089	NS
Residual	1358.8218	3288.7985	8937.1166	977.1358	12	81.4280	
Treatment	161.7795	450.7005	1467.6755	555.7806	4	138.9452	17.3788**
IQxTreatment	15.0529	5.8412	66.5440	64.8602	4	16.2151	2.0281^{NS}
Residual	128.6785	76.6929	453.4571	407.7478	51	7.9951	

*$p < .05$, **$p < .01$

TABLE V

Comparison of Mean Scores in Experiments A, B, and C
Responses to the Post-test

Treatment Means

Experiment A	$\overline{X}_{E_i}$	$\overline{X}_{E_i} - \overline{X}_{E_3}$	$\overline{X}_{E_i} - \overline{X}_{E_2}$	$\overline{X}_{E_i} - \overline{X}_{E_1}$	
$\overline{X}_{E_5}$	25.0649			1.5966	
$\overline{X}_{E_1}$	23.4683	5.1424**	-.2059		
$\overline{X}_{E_2}$	23.6742	5.3483**			
$\overline{X}_{E_3}$	18.3259				
Experiment B					
$\overline{X}_{E_5}$	17.0958			4.2948	
$\overline{X}_{E_1}$	12.8010	1.3325	-7.4337**		
$\overline{X}_{E_2}$	20.2347	8.7662**			
$\overline{X}_{E_3}$	11.4685				
Experiment C	$\overline{X}_{E_i}$	$\overline{X}_{E_i} - \overline{X}_{E_4}$	$\overline{X}_{E_i} - \overline{X}_{E_3}$	$\overline{X}_{E_i} - \overline{X}_{E_2}$	$\overline{X}_{E_i} - \overline{X}_{E_1}$
$\overline{X}_{E_6}$	14.2591				.0565
$\overline{X}_{E_1}$	14.2026	9.3298**	4.6363**	1.0640	
$\overline{X}_{E_2}$	13.1386	8.2658**	2.5723		
$\overline{X}_{E_3}$	9.5663	4.6935**			
$\overline{X}_{E_4}$	4.8727				

$^*p < .05$, $^{**}p < .01$

E_1 and E_3, with program E_1 producing the larger mean.

There were no significant differences as a result of variations in the number of associations of attributes *per concept* in any of the three experiments.

Since it could be argued that achievement may have been a function of the time spent on completing the program, an analysis of variance was performed, for Experiments A and B, to test the following two hypotheses: (1) efficiency, as measured by the ratio of the score obtained on the criterion items, for the respective programs, to the time required to complete the respective programs, is a function of (a) the number of associations of attributes *per frame*, (b) the level of student ability, or (c) the interaction of (a) and (b); (2) efficiency, as measured by the ratio of the achievement of the *S* on the criterion items to the amount of time required to present the frame is a function of (a) the number of associations of attributes *per concept*, (b) the level of student ability, or (c) the interaction of (a) and (b).

The results of the analyses for each of the experiments are presented in Table VI.

TABLE VI

Analyses of Variance of the Efficiency Ratios for Experiments A and B

Experiment A

Source	SS	df	MS	F
I.Q.	4.8885	1	4.8885	6.6186*
Residual	9.6018	13	.7386	
Treatment	7.5175	3	2.5058	51.1388**
IQxTreatment	.3696	3	.1232	2.5143^{NS}
Residual	1.9103	39	.0490	
Experiment B				
I.Q.	1.6865	1	1.6865	1.1419^{NS}
Residual	11.8152	8	1.4769	
Treatment	8.4131	3	2.8044	37.0952**
IQxTreatment	.1956	3	.0652	NS
Residual	1.8146	24	.0756	

*$p < .05$

**$p < .01$

Treatment effects were significant in both Experiments A and B (Experiment A, F = 51.1888, df = 3/39; Experiment B, F = 37.0952, df = 3/24). Level of ability effects were significant in Experiment A (F = 6.6186, df = 1/13), with the high I.Q. group performing significantly better than the low I.Q. group. The means were compared using Tuckey's Method for Multiple Comparisons; the same comparisons that were made with the criterion items were made with the means of efficiency. The results of the comparisons for Experiments A and B appear in Table VII. The Ss were not timed in Experiment C and therefore the efficiency scores could not be calculated.

TABLE VII

Comparison of Mean Scores on Efficiency

Experiment A				
	$\bar{X}_1$	$\bar{X}_1 - \bar{X}_3$	$\bar{X}_1 - \bar{X}_4$	$\bar{X}_1 - \bar{X}_6$
$\bar{X}_8$	1.4749			-.9184**
$\bar{X}_6$	2.3933	.5290**	.1907	
$\bar{X}_4$	2.2026	.3383**		
$\bar{X}_3$	1.8643			
Experiment B				
$\bar{X}_8$	1.6151			-.9285**
$\bar{X}_6$	2.5436	-.3120	.0992	
$\bar{X}_4$	2.4444	-.4412**		
$\bar{X}_3$	2.8556			

**p < .01

In Experiment A, there were significant differences in efficiency as a result of variation in the number of associations of attributes per frame. It can be observed that the means for programs E_1 and E_2 were significantly greater than the mean for program E_3. The comparison of the means of programs E_1 and E_2 revealed no significant differences.

In Experiment B, the only comparison that was significant was a comparison of programs E_2 and E_3.

In both Experiments A and B, there were significant differences as a function of the number of associations of attributes per concept. The same two comparisons were made in each case, and both were significant; i.e., the comparison of the means for programs E_1 and E_5.

An analysis of variance was computed to test the hypotheses that (1) the subjective difficulty ratings of the concept by S are a function of (a) the number of associations of attributes per frame, (b) the level of student ability, or (c) the interaction of (a) and (b); (2) the subjective difficulty ratings of S are a function of the (a) number of associations of attributes per concept, (b) the level of student ability, or (c) the interaction of (a) and (b). The ratings were transformed into numerical data by assigning the value of one to a response of very easy, two to a response of easy, etc. The subjective difficulty ratings of the concepts were then summed over the individual items in the rating list under each of the experimental conditions, so that there was again one score under each experimental condition for each S. These sums were used in the analyses.

The results of the analyses for each of the experiments are presented in Table VIII.

TABLE VIII

Analyses of Variance of the Subjective Difficulty Ratings of the Concepts - Experiments A, B, and C

Experiment A				
Source	SS	df	MS	F
I.Q.	22.0828	1	22.0828	NS
Residual	653.0022	13	50.2309	
Treatment	358.6883	3	119.5628	23.5253
IQxTreatment	7.0628	3	2.3543	NS
Residual	198.2103	39	5.0823	
Experiment B				
I.Q.	12.1000	1	12.1000	NS
Residual	1533.8000	8	191.7250	
Treatment	292.4000	3	97.4667	14.4932
IQxTreatment	10.7000	3	3.5667	NS
Residual	161.4000	24	6.7250	
Experiment C				
I.Q.	37.1491	1	37.1491	NS
Residual	1042.3179	13	80.1783	
Treatment	613.7799	4	153.4450	35.6062**
IQxTreatment	59.9068	4	14.9767	3.4753*
Residual	224.0928	52	4.3095	

$^*p < .05$, $^{**}p < .01$

Differences among the experimental treatments were significant in all three experiments (Experiment A, F = 23.5253, df = 3/39; Experiment B, F = 14.4932, df = 3/24; Experiment C, F = 35.6062, df = 4/52). Differences between levels of ability were not significant. The interaction of level of ability and treatments was significant only in Experiment C (F = 3.4753, df = 4/52). The means were compared for each of the experimental conditions using Tuckey's Method for Multiple Comparisons; the same comparisons that were completed for the criterion items were performed with these means. The results appear in Table IX.

TABLE IX

Comparison of Mean Scores on Subjective Difficulty Ratings

	$\overline{X}_i$	$\overline{X}_i - \overline{X}_3$	$\overline{X}_i - \overline{X}_4$	$\overline{X}_i - \overline{X}_6$
Experiment A				
$\overline{X}_8$	31.37			2.54**
$\overline{X}_6$	28.83	-6.62**	-1.64	
$\overline{X}_4$	30.47	-4.98**		
$\overline{X}_3$	35.45			
Experiment B				
$\overline{X}_8$	30.90			-4.0*
$\overline{X}_6$	34.90	-2.80	3.20*	
$\overline{X}_4$	31.70	-6.00**		
$\overline{X}_3$	37.70			

*p $<$ .05, **p $<$ .01

	$\overline{X}_i$	$\overline{X}_i - \overline{X}_1$	$\overline{X}_i - \overline{X}_3$	$\overline{X}_i - \overline{X}_4$	$\overline{X}_i - \overline{X}_6$
Experiment C (Treatment Means)					
$\overline{X}_{10}$	37.07				1.34
$\overline{X}_6$	30.73	-6.54**	-5.80**	0.00	
$\overline{X}_4$	30.73	-6.54**	-5.80**		
$\overline{X}_3$	36.53	.74			
$\overline{X}_1$	37.27				
Experiment C (Interaction)					
High I.Q. $\overline{X}_{10}$	32.38				-.50
$\overline{X}_6$	32.88	-5.38**	-3.63**	2.25*	
$\overline{X}_4$	30.63	-7.63**	-5.88**		
$\overline{X}_3$	36.50	-1.75			
$\overline{X}_1$	38.25				
Low I.Q. $\overline{X}_{10}$	31.71				3.43**
$\overline{X}_6$	28.29	-7.86**	-8.29**	-2.57*	
$\overline{X}_4$	30.86	-5.29**	-5.71**		
$\overline{X}_3$	36.57	.43			
$\overline{X}_1$	36.14				

*p $<$ .05, **p $<$ .01

To determine whether the variations in the number of associations per frame produced a statistically significant variation in the achievement of the learner, the following means were compared in Experiment A: the means of programs E_1, E_2, and E_3. The comparisons of the means for programs E_1 and E_2 reveal that the concepts within those programs were rated significantly easier than the concepts within program E_3. There were no significant differences in rating between the means of programs E_1 and E_2.

In Experiment B, the concepts in program E_1 were considered to be significantly more difficult than the concepts in program E_2. There were significant differences between the means of programs E_1 and E_3, with the concepts in program E_3 receiving a higher rating of difficulty than the concepts in program E_2. The contrast between the means of programs E_1 and E_3 was not significant.

In Experiment C, the concepts in programs E_1 and E_2 were considered to be significantly less difficult than the concepts in programs E_3 and E_4. The contrasts between the means of programs E_3 and E_4 and the contrasts between the means of programs E_1 and E_2 were not significant.

To determine whether the variations in the number of associations per concept produced a statistically significant variation in Ss subjective difficulty ratings, the means for programs E_1 and E_5 were compared in Experiments A and B. For Experiment A, the concepts in program E_1 were rated significantly easier than those in program E_5; in Experiment B, the opposite was true. The contrast of the means of programs E_1 and E_6 was not significant in Experiment C.

To determine whether (1) the interaction of the level of student ability and the number of associations of attributes per frame and/or (2) the interaction of the level of student ability and the number of associations of attributes per concept produced a statistically significant variation in Ss subjective difficulty ratings, the means were first calculated for each level of ability. Comparisons were then made within those levels.

The testing of the first hypothesis involved comparisons of the means of programs E_1, E_2, E_3 and E_4 in both ability groups. For the high ability group, the concepts in programs E_1 and E_2 were rated significantly easier than the concepts in programs E_3 and E_4. The mean of program E_1 was significantly greater than the mean of program E_2, indicating that the concepts in program E_1 were considered to be significantly easier than the concepts in program E_2. The contrast between the means of programs E_3 and E_4 was not significant.

For the low ability group, the concepts in programs E_1 and E_2 were rated significantly easier than the concepts in programs E_3 and E_4. The mean of program E_1 was significantly greater than the mean of program E_2, indicating that the concepts in that program were considered to be significantly more difficult than the concepts in the E_2 program. The contrast between the means of programs E_3 and E_4 was not significant.

The testing of the second hypothesis involved comparing the means of programs E_1 and E_6 in both ability groups. In the high ability group, there was no significant difference between the two means. However, in Experiment B, the concepts in program E_1 were considered to be significantly easier than the concepts in program E_6.

DISCUSSION

In the preceding section, statistical evidence was presented to support the hypothesis that achievement is a function of the number of associations of attributes per frame. It was observed that there was a high degree of consistency in the results of the analyses of the criterion items and post-tests, for all three experiments.

Specifically, although the differences were not significant, the mean criterion item scores of program E_2 were greater than the mean criterion item scores of program E_1 which presented the six attributes in six frames in both Experiments A and B. Furthermore, the post-test mean of program E_2 was greater than the mean for program E_1 in the first experiment and was significantly greater than the mean of program E_1 in Experiment B. The mean of program E_2 also was significantly greater on the criterion items than the mean of program E_3 which presented the six attributes in three frames in Experiment A and was greater, though not significantly, than the mean of program E_3 in Experiment B. On the post-test, the mean of program E_2 also was significantly greater than the mean of program E_3 in Experiments A and B. In Experiment C, the mean for the post-test of program E_2 was lower than the mean of program E_1, although not significantly. The mean of program E_4, which presented six attributes in one frame, also was significantly less than the means of programs E_1, E_2, and E_3 in Experiment C.

Although support for the hypothesis was gained in the analysis of the responses to the criterion items in Experiment C; i.e., the means of criterion items differed significantly as a function of the number of associations per frame, there was an inconsistency in the direction of the differences between means. Specifically, the mean of program E_4 was significantly greater than the mean of program E_2, which presented the six attributes in four frames. An explanation for this lack of directional consistency may be that Ss read the answers to the criterion items before writing their responses[1]; it was not possible for Ss to read the answers to the criterion items in the first two experiments. The fact that the mean achievement of Ss on the criterion items for the respective programs was consistently higher in Experiment C than in either Experiment A or B would substantiate the hypothesis that Ss read many of the answers to the criterion items before responding.

The results of this study tend to support the findings of Coulson and Silberman (1960) and Campbell (1961). On the other hand, the results also tend to refute their findings. When one compares the means of program E_1 and program E_2 with those for program E_3, one will note that the results are in agreement with those of Coulson and Silberman (1960) and Campbell (1961); i.e., the small-step programs were superior to the large-step programs. However, the comparison of the means for program E_2 with those for program E_1 tends to refute their findings; i.e., the median program is superior to the small-step program. An explanation of the apparent inconsistencies of the present study with previous ones may be that while a small-step program may be more effective than a large-step program, a median step program may be more effective than either a small- or large-step program.

---

[1] This may have occurred in spite of instruction to the contrary.

One may infer from the results of this study that achievement is affected by step-size which can be determined adequately by the number of associations of attributes per frame.

In addition to the evidence supporting the hypothesis that achievement is a function of the number of associations of attributes per frame, it was observed that the presentation of too many or too few attributes per frame reduced the effectiveness of the instructional process. The fact that the mean achievement of S̲s on the post-test of Experiment B for program E_2 was greater than their achievement on the post-test for program E_1, and the fact that the mean achievement for program E_2 was significantly greater than the mean for program E_4 on the post-test for S̲s on all three experiments, would support this observation.

The results of this study are consistent with the findings of Evans, Glaser and Homme (1960). They compared performance of four groups on retention tests and found that S̲s did better on the very small-step forms (51 to 68 steps) than on the very large-step forms (30 and 40 steps). However, they also found that the intermediate step-size (51 steps) provided better achievement, although not significantly so, than the very small-step size (68 steps). Note that the results of this study are consistent with this general pattern. An explanation for these results may be (1) that the presentation of one attribute per frame in the six frame program (E_1) afforded little opportunity for the association among attributes of the concept to be learned, thus reducing achievement, and (2) that there are limitations on the number of associations of attributes per frame that a student may profitably absorb before responding to a criterion item. These findings are consistent with Miller (1956) and Jensen (1962) who found that there are limitations to the amount of information that can be absorbed as a single event. One could conclude that an optimal number of associations per frame is required for achievement to be maximized.

An explanation for the failure to obtain significant differences in achievement as a function of level of ability or the interaction of student ability and the number of associations of attributes per frame, may be that the range of I.Q.'s was too small. In none of the three experiments was the difference in mean I.Q. between the high and low ability groups greater than 17.39 points.

The results of these experiments support the studies of Maccoby and Sheffield (1958) and of Margolius and Sheffield (1961). Both of these investigations concluded that a procedure which began with short steps and gradually increased step-size was superior to a self-pacing method in which the student set his own step-size. It will be noted that in Experiment A the mean achievement of Ss on the criterion items in program E_1 was greater than their achievement on the criterion items in program E_3, while in Experiment B the opposite was true. The S̲s employed in the first experiment were naive with respect to the experimental materials while S̲s in Experiment B had experience with the programs in Experiment A. Thus, conditions were provided for the occurrence of this phenomenon.

The hypothesis that achievement is a function of the number of associations of attributes per concept is also substantiated by the results of the three experiments. The mean of program E_5 (twelve associations of attributes per concept) was significantly greater than the mean of the E_1 program (six associations of attributes per concept) for the criterion items in Experiments A and B. In Experiment C, the mean of program E_6 (eighteen associations of attributes per concept) was greater than the mean of program E_1, although the differences were not significant. The post-test scores for programs E_5 and E_6 were greater than those for program E_1. Thus, it may be concluded that these results indicate that as the number of associations of attributes per concept increases, the achievement of the learner increases.

This finding supports the work of Campbell (1961). He found that S̲s using a small-step program did significantly better than S̲s using a large-step program; there is a similarity between the materials employed by Campbell (1961) and the materials employed to investigate the hypothesis in this study. In Campbell's (1961) study, the large-step form differed from the short-step form only in the absence of remedial loops; thus, it could be argued that S̲s received more associations of attributes per concept in the small-step form than they received in the large-step form.

Since support for the hypothesis that achievement is a function of the number of associations of attributes per frame and of the number of associations of attributes per concept has been provided by the results of this study, it appears probable that the interaction of the number of associations per frame and the number of associations per concept might result in greater acquisition than either variable by itself. Consideration of (1) the fact that the mean of program E_2 was greater or significantly greater than the mean of program E_1 in Experiment A or B, and (2) the fact that the means of programs E_5 and E_6 were greater or significantly greater in all of the analyses, supports the hypothesis that there may be an interaction of the two variables.

The hypothesis that efficiency, as measured by the ratio of the score on the criterion items for the respective programs to the time to complete the respective programs, is a function of the number of associations of attributes per frame and the level of student ability, was supported by the results of Experiments A and B.

In Experiment A, program E_1 was significantly more efficient than program E_3, while in Experiment B program E_1 was less efficient than program E_3, although the difference was not significant. There were no significant differences between the means of programs E_1 and E_2. In Experiment A, program E_2 was significantly more efficient than program E_3, while in Experiment B, the opposite was true. These conflicting results suggest that the experiences of Ss in working with programmed materials may have enabled them to associate a greater number of attributes per frame than they had in Experiment A. This view is substantiated by the mean achievement of Ss on the criterion items in the respective programs. As pointed out earlier in the discussion, these results tend to support the work of Maccoby and Sheffield (1958) and of Margolius and Sheffield (1961) who found that a procedure which began with short steps and gradually increased step-size was superior to a self-pacing method in which the student set his own step-size.

Experiments A and B also provide support for the hypothesis that efficiency is a function of the number of associations of attributes per concept. In both experiments, program E_1 (six associations of attributes per concept) was more efficient than program E_5 (twelve associations of attributes per concept). An explanation for these results may lie in the fact that Ss required more time to make the twelve responses in program E_5 than they required to make the six responses in program E_1. This explanation supports the study of Campbell (1961), who found that small-step forms took more time to complete than did large-step forms.

Partial support was provided for the hypothesis that the subjective difficulty ratings of the concepts by Ss is a function of the number of associations of attributes per frame, the level of student ability, or the interaction of these two variables. The results show that the mean for program E_3 was significantly greater than the mean for program E_2 in all three experiments. The mean for program E_3 also was significantly greater than the mean for program E_1 in Experiments A and C, and was greater, although the differences were not significant, in Experiment B. In Experiment A the mean for program E_2 was greater than the mean for program E_1; in Experiment B, the opposite was true; there were no significant differences between the two means in Experiment C. The programs which were difficult for Ss, as measured by the students' achievement, were also perceived as being difficult. The mean for program E_4 which presented the six attributes in one frame, was significantly greater than the mean for program E_1, and program E_2 in Experiment C, and was greater than the mean for program E_3. One may infer that the subjective difficulty ratings were a function of the number of associations per frame. It would appear that Ss rated the concepts in relation to their achievement on the respective programs; i.e., the higher the achievement of S, the less difficult he thought the concept to be. Support for this hypothesis may be found in the observations that there is a close correspondence between the achievement of Ss, both on the criterion items and the post-tests, and Ss'ratings. The effect of student ability upon the ratings was not significant in any of the experiments; an explanation may be that the small I.Q. range was not great enough to provide a significant difference.

In Experiment C there was a significant interaction of I.Q. and treatments. Inspection of the table of means (Table IX) provides support for the hypothesis that the subjective difficulty ratings are a function of the interaction of the number of associations of attributes per frame and the level of student ability. High ability Ss considered the concepts in program E_1

to be significantly more difficult than the concepts in program E_2; for the low ability Ss the opposite was true. An explanation for these results may be that high ability Ss might be able to associate a greater number of attributes per frame than low ability Ss; consequently, they considered the concepts in program E_2 to be less difficult than the concepts in program E_1 since their achievement was higher on program E_2. Low ability Ss considered the concepts in program E_1 to be less difficult than those in program E_2. An inspection of the mean achievements of Ss on the post-test in Experiment C provides support for this hypothesis. For high ability Ss, the mean achievement on program E_2 was greater than the mean achievement on program E_1; the reverse was true for the low ability Ss. Consideration of this fact suggests that varying the number of associations of attributes per frame may be a meaningful technique to provide for individual differences among learners.

The effect of the number of associations of attributes per concept upon the ratings of the concepts by Ss was significant in two of the three experiments. The concepts in program E_5 were considered to be more difficult than those in program E_1 in Experiment A; the opposite was true in Experiment B. The concepts in program E_6 were considered to be more difficult than the concepts in program E_1 in Experiment C. An explanation for these results may be that an interaction of the number of associations per frame and the number of associations of attributes per concept affected the difficulty ratings.

Inspection of the table of means (Table IX) in Experiment C reveals that support is provided for the hypothesis that the ratings are a function of the interaction of the level of student ability and the number of associations of attributes *per concept*. The low ability Ss considered the concepts in program E_6 to be more difficult than the concepts in program E_1, while the high ability Ss found no significant differences between them. These results would suggest that high ability Ss did not consider an increase in the number of associations of attributes per concept as an increase in the difficulty of that concept, while the low ability Ss found such an increase in difficulty.

From the findings of this study it may be concluded that it is possible to determine the optimal number of associations per frame and the number of associations per concept required to teach a concept of a known number of attributes. It also may be concluded that the efficiency of the program is affected by the number of associations of attributes per frame, the number of associations of attributes per concept, and the level of student ability. Further, it may be concluded that Ss rated the concepts in relation to their achievement on the respective programs.

EXPERIMENT III

Introduction

In Experiment II, it was observed that the number of attributes associated per frame and the number of associations of attributes per concept affected learner acquisition in the association stage of concept learning. This more precise definition of program development may explain in part the failure to obtain differences in Ss' achievement as a function of size-of-step in Experiment I.

Some preliminary data was secured in Experiment II which suggest another revision in programming technique, one which may have confounded the results in Experiment I. Specifically, the higher achievement of Ss on those concepts which were presented in eight frames, two of which were practice frames[1], presented during learning of the concepts, suggests that practice during or prior to the learning of a concept may be a means of varying step-size.

Because of the failure to include practice in the association stages of concept learning in Experiment I with the resulting failure to obtain differences in achievement, and the tentative findings in Experiment II of the effects of practice on learner performance in the association stage of concept learning, it seemed of some importance to determine systematically the effects of practice on learner acquisition of the association stage of learning a concept.

---

[1] Frames which required responses which had been learned previously.

Objective: The purpose of this experiment was to determine the effects of practice on the association stage of concept learning for Ss of two levels of ability.

METHOD

Subjects

The Ss were 116 eighth-grade students enrolled in four general science classes.[1] Using the median I.Q. score from the California Test of Mental Maturity, Ss were classified as "high" (above median) or "low" (below median) on ability. The "high" ability Ss were stratified on the basis of sex and randomly assigned to one of two groups. The same procedure was repeated for the "low" ability group. Since two teachers were involved, equal numbers of Ss from each class were assigned to each group. The range of I.Q.'s for the "high" group was 111 to 139 and for the "low" group, 83 to 111, with a mean of 118 for the "high" group and a mean of 96 for the "low" group.

Materials

Seven brief programs on biological classification were developed for use in the experiment. Each program contained six characteristics (attributes of the concept) useful in identifying a class of animals. The program for one class of animals will be found in Appendix C.

To control for programming techniques, each teaching frame presented two attributes (characteristics) and S was asked to respond to a syllogism. Each teaching frame was followed by a practice item which required practice of all characteristics of the class presented to that point; i.e., first, two characteristics were practiced, then four, and finally six. The frames in which practice of all six characteristics was required were used as post-test items.

---

[1] The Ss were eighth-grade students in the Sunbury Junior-Senior High School, Sunbury, Pennsylvania.

The programmed frames were prepared on transparencies for use with an overhead projector. The projector was equipped with a masking device which allowed the operator to present one frame at a time. From a pilot study it was found that the optimal exposure time for each type of frame was: teaching frames, 50 seconds; answer frames, 10 seconds; and practice frames, 40 seconds, with the exception of the last practice frame which was used as a post-test question and which was presented for 75 seconds.

An additional question was included as the last frame for each class of animals (Appendix C, #7). This item was used to determine Ss' ability to apply the knowledge acquired; i.e., to determine whether the information which S had acquired could be used meaningfully or only by rote. This frame was exposed for 50 seconds.

Two forms of the program were prepared, a practice version and a no-practice version. In the no-practice form, frame 2, which permitted practice on two of the six characteristics of a class, and frame 4, which permitted practice on four of the six characteristics of a class, were omitted.

Procedure

Each ability group was randomly assigned to a "practice" and a "no-practice" condition. The experimental design used is presented in Table I.

TABLE I

Experimental Design

	Practice	No practice
High		
Low		

A pretest (Appendix B) consisting of 16 items was administered to all Ss. The following day separate sessions were arranged for the groups to begin work on the programs. The E used one program on one of the seven classes of animals to familiarize Ss with the task. The Ss were required to write their answers on prepared sheets and place the completed sheets in a box containing an aperture through which the answer sheets could be slid. Each S placed his answer to each frame into a box prior to the exposure of the correct answer for a frame. No answers were shown to frames 6 and 7 of each unit. After completing the practice program, Ss worked through three of the programmed units. On the next day, they completed the remaining three units. The order of presentation of the units was the same for all groups. Table II includes the names of the units used and their order of presentation.

TABLE II

Titles and Presentation Order of Experimental Units

Session I	Session II
Introduction - Class Amphibia	
I Class Asteroidea	IV Class Arachnoidea
II Class Cephalopodea	V Class Mammalia
III Class Aves	VI Class Reptilia

The answer sheets for the six units, including Ss' responses to items 6 and 7, were collected and scored. A total score for each S was obtained by summing the correct number of responses to item 6 on each of the programs. A similar procedure was used for obtaining a total score for item 7.

RESULTS

In order to determine the effect of practice on achievement, a two-way analysis of variance was computed on the scores obtained from the criterion frames (Item 6 in each unit). Table III presents a summary of this analysis. Table IV includes the mean scores for each group on the six criterion frames.

TABLE III

Analysis of Variance of Criterion Frame Scores (Item 6)

Source	SS	df	M.S.	F	P
Ability	1265.17	1	804.02	30.56	<.01
Practice	804.02	1	1265.17	48.09	<.01
error	2631.02	100	26.31		
Total	4700.21	102			

TABLE IV

Mean Scores on Criterion Frame. (Maximum score = 36)

Ability	Practice	No practice	Total
High	33.32 N = 25	29.79 N = 24	31.59
Low	28.21 N = 29	20.36 N = 25	24.57
Total	30.57	24.98	27.91

From Table III, it is seen that the F's for ability and for practice reached significance. The high ability Ss scored significantly higher than the low ability Ss and both practice groups were higher in achievement than the no-practice groups.

A similar analysis was completed on the scores obtained for the application frames (Item 7 in each unit). Table V summarizes this analysis, and Table VI presents the means.

TABLE V

Analysis of Variance of Application Frame Scores (Item 7)

Source	SS	df	M.S.	F	P
Ability	69.847	1	69.847	24.177	<.01
Practice	.839	1	.839	.29	>.05
Error	288.868	100	2.889		
Total	359.554	102			

TABLE VI

Mean Scores on Application Frame. (Maximum score = 12)

Ability	Practice	No practice	Total
High	11.36	11.46	11.41
Low	10.00	9.48	9.76
Total	10.62	10.44	10.54

The results in Table V indicate that the groups differed in achievement as a function of ability, with the high ability Ss scoring higher on the application frames. However, neither the high nor the low ability group differed in achievement as a function of practice.

DISCUSSION

The results of this study support the finding of Experiment II, that practice of attributes during the association stage of learning a concept does increase achievement significantly at this stage of concept learning. If the findings of this study can be compared to studies of "whole vs. part learning", it would suggest that the practice condition in the experiment may be in fact a "part-whole" method of learning, while the no-practice during learning may be a "whole" method of learning. These findings would tend to support the "part-whole" method of learning in the association stage of concept learning.

It was of some interest to note that both ability groups achieved significantly more with practice than without. The difference in achievement as a function of practice was almost twice as great for the high I.Q. group as for the low I.Q. group.

As would be expected, the high ability Ss, other things being equal, required less practice to achieve the same level of competency during the association stages of learning than do low ability Ss.

Assuming that acquisition of attributes has taken place prior to practice, it may also be concluded from this study that practice is of importance (1) during the association stages of learning, and (2) following conceptualization, if the concept is to be useful.

These findings would tend to offer additional explanation why the programming technique used in Experiment I was not adequate in difficulty between difficult and easy versions of the instructional programs.

An analysis of the criterion items which evaluated the Ss' ability to apply the information learned in acquisition frames yielded no significant differences as a function of the independent variable. This failure to obtain differences probably resulted from a failure of the test to discriminate. It is interesting to observe the high scores obtained on these items. This suggests that the Ss did not learn exclusively in a rote fashion.

EXPERIMENT IV

Introduction

The findings of Experiments II and III would suggest that the following be considered as a more adequate basis for varying size-of-step in programming the procedures used in Experiment I:

(1) a technique of programming which is based on the association of attributes in the learning of a concept,

(2) the number of attributes to be associated per frame, and

(3) the systematic presentation of practice frames during the association stages of concept learning.

It was the purpose of this experiment to evaluate the application of this method of programming to an organized body of content.

Objective: To determine the effects of varying the number of attributes associated per frame, varying the amount of practice of attributes prior to conceptualization, and the interaction of these two variables on Ss' achievement and their perceived concept difficulty.

METHOD

Subjects

The Ss were 120 eleventh-grade students enrolled in two classes in health.[1] Using the median gamma score of the Otis

---

[1] The Ss were students enrolled in the eleventh grade of Western Area Joint High School, Mifflinburg, Pa.

Mental Ability test, the Ss were classified as "high" ability (above median) or "low" ability (below median). The mean I.Q. for the high ability group was 113, with a range of 104-129, and for the low ability group the mean I.Q. was 94, with a range of 73-103. The high group was stratified on the basis of sex and randomly assigned to one of four groups. The low group was assigned to one of the four groups in a similar manner. Since there was a different teacher for each class, equal numbers of Ss from each class were assigned to each group.

Materials

A program (121 frames) on the elementary concepts of maturation and growth was developed. The subject matter was selected on the assumption that Ss would be relatively unfamiliar with the topic, yet the topic would add relevant information to the Ss' knowledge of health. The program was a constructed response, linear type. All teaching items, practice items, and criterion items were clearly identified. As a rule, one attribute was presented per frame, although on occasion, two attributes were presented.

Three additional forms of the program were developed. In order to determine the effects of practice on achievement, Form II of the program was constructed by removing the practice items from Form I (the original program). In order to investigate the effect of presenting a greater number of attributes per frame, Form III presented twice the number of attributes in each frame as Forms I and II, resulting in a shorter program.

To determine the effects of practice on a program containing a larger number of attributes per frame, Form IV was developed by removing the practice items from Form III. No confirmation was provided in any of the forms. Examples of the programs are presented in Appendix D.

The experimental groups were randomly assigned to the programs. Table I presents the experimental design.

TABLE I

Experimental Design

Ability	(long) Form I 121 frames with practice	(long) Form II 94 frames without practice	(short) Form III 95 frames with practice	(short) Form IV 68 frames without practice
High				
Low				

Procedure

The pretest (Appendix D) containing twenty-four items which were included in the program as criterion items, was administered to all Ss in one session. Following the test, E instructed Ss on the use of the program, asked them to write their responses on the program booklet, and informed them that errors on the program would not be used to adjust their score on the final test. Since E monitored the work sessions, and the teachers did not participate in the instruction of the material, the teacher variable was eliminated. The classes worked 40 minutes daily during their regular class period. As each S completed the program, E collected the material, tabulated the time which each needed to complete the program and administered a questionnaire (Appendix D). Each S received the post-test during the first class period following his completion of the assignment. This procedure was continued until all Ss had completed the program, the perceived difficulty questionnaire, and the post-test. The pretest and post-test were identical.

RESULTS

Post-Test Achievement

In order to investigate the effects of practice, the amount of information per frame, and the interaction of these two conditions on achievement, a two-way analysis of covariance, using pretest scores as control variables, was computed on the post-test scores for all experimental groups. Table II presents a summary of this analysis. Table III presents the adjusted mean post-test scores.

TABLE II

Analysis of Covariance of Post-Test Scores

Source	SS	df	M.S.	F	P
Ability	262.22	1	262.22	23.13	$< .01$
Program	81.26	3	27.09	2.39	$> .05$
Interaction	219.84	3	73.28	6.46	$< .01$
Error	1111.06	98	11.34		
Total	1674.38	105			

It can be observed from Table II that significant *F*'s were obtained for ability and interaction.

TABLE III

Adjusted Mean Post-Test Scores

	Form I	Form II without practice	Form III	Form IV without practice	Totals
High	11.84 N = 12	9.37 N = 11	9.86 N = 15	10.14 N = 15	10.28
Low	7.05 N = 12	8.56 N = 13	5.26 N = 14	5.79 N = 15	6.59
Totals	9.44	7.64	7.96		

A *t*-test analysis of the "high" groups revealed that Form I (long with practice) resulted in significantly higher scores, $t = 2.466$, $df = 21$, $p < .03$, than Form II (long without practice). A *t*-test analysis of the "low" groups indicated that Form II (long without practice) yielded significantly higher scores, $t = 2.773$, $df = 26$, $p < .01$, than Form IV (short without practice). Form I (long with practice) tended to result in higher post-test achievement, $t = 1.789$, $df = 24$, $p < .08$, than Form III (short with practice).

The significant *F* for ability (Table II) indicated that in general the "high" group scored significantly higher than the "low" group.

Time

In order to determine whether the programs produced differences in working time, a two-way analysis of variance was computed on the time needed by each S to finish the program. Table IV summarizes this analysis. Table V presents the mean time in minutes used to complete the program by each group.

TABLE IV

Analysis of Variance of Time Required to Complete the Experimental Program

Source	SS	df	M.S.	F	P
Ability	1910.37	1	1910.37	12.76	(<.01)
Program	4751.83	3	1583.94	10.58	(<.01)
Interaction	115.25	3	38.42	.26	(>.05)
Errors	14671.61	98	149.71		
Total	21449.06	105			

TABLE V

Mean Time Required to Complete the Four E Programs

	Form I	Form II without practice	Form III	Form IV without practice	Total
High	69.17 N = 12	55.45 N = 11	65.33 N = 15	53.00 N = 15	60.66
Low	79.17 N = 12	62.69 N = 13	72.86 N = 14	62.86 N = 15	69.15
Total	74.17	59.38	68.97	57.76	64.90

It can be observed in Table IV that significant F's for ability and programs were obtained. In general, the "high" group required significantly less time than the "low" group to complete the experimental materials.

A t-test analysis comparing time taken on the various programs resulted in the following: Form I required significantly more time to complete than Form II ($t = 4.40$, $df = 46$, $p < .01$) and Form IV ($t = 5.488$, $df = 51$, $p < .01$); and Form III required significantly more time to complete than Form II ($t = 2.641$, $df = 51$, $p < .02$) and Form IV ($t = 3.460$, $df = 56$, $p < .01$). These results indicate that the removal of practice frames results in significantly less time to complete the program; however, a reduction in the length of the program by presenting more information per frame does not require less time to complete.

Since both programs with practice required the same number of responses as did both programs without practice[1], a more meaningful way of looking at this variable may be through an analysis of the amount of time needed for each response.

Table VI presents the mean number of responses completed by each group in an average 10-minute working period.

TABLE VI

Mean Number of Responses Completed by E Groups in a 10-minute Period

	Form I	Form II	Form III	Form IV	Total
High	3.13	3.34	3.01	3.25	3.16
Low	3.58	3.78	3.36	3.86	3.61
Total	3.36	3.58	3.18	3.54	3.38

---

[1] Programs differed only in the number of attributes presented per frame; thus, the total number of frames differed, while the total number of responses did not.

Table VI indicates, in general, that the removal of practice items increases the time per response. It is interesting to observe, also, that the low ability group spent less time making each response than the high ability group.

Efficiency

To determine the relationship of time and achievement, efficiency scores were computed by dividing each post-test achievement score into the time required to complete the program. The lower the efficiency score, the more efficient the time spent on the program. A two-way analysis of variance using efficiency scores as a dependent variable to determine the effects of ability, the E programs, and the interaction of these two variables, was performed. Table VII summarizes this analysis and Table VIII presents the mean efficiency scores for each group.

TABLE VII

Analysis of Variance of Efficiency Scores

Source	SS	df	M.S.	F	P
Ability	910.39	1	910.39	12.02	<.01
Program	471.46	3	57.15	2.08	>.05
Interaction	328.38	3	109.46	1.44	>.05
Error	7344.30	97	75.71		
Total	9054.53	104			

Table VII indicates that the "high" group was more efficient in their use of time than the "low" group. The F's for program and for interaction indicate no differences.

TABLE VIII

Mean Efficiency Scores

	Form I	Form II	Form III	Form IV	Total
High	4.52	4.51	5.56	3.93	4.65
Low	10.66	6.01	16.31	8.77	10.54
Total	7.59	5.30	10.75	4.54	7.56

Questionnaire

In order to analyze perceived difficulty as influenced by ability and programs, a two-way analysis of variance was computed on the questionnaire scores. These scores were obtained by assigning numerical values to each choice. A response of very easy received 1 point and a response of very difficult, 5 points. The scores for each scale were summed for each S. Table IX presents the summary of the analysis. Table X presents the mean scores for each group.

TABLE IX

Analysis of Variance of Difficulty Estimates

Source	SS	df	M.S.	F	P
Ability	183.96	1	183.96	15.89	$<.01$
Programs	63.79	3	21.26	1.84	$>.05$
Interaction	124.54	3	41.51	3.58	$<.03$
Error	1099.81	95	11.58		
Total	1472.10	102			

Table IX indicates that the perceived difficulty scores were differentially influenced by ability and the interaction of ability and programs. The high group perceived the concepts as easier than the low group.

TABLE X

Mean Perceived Difficulty Scores

	Form I	Form II without practice	Form III	Form IV without practice	Totals
High	14.33 N = 12	14.72 N = 11	16.13 N = 15	14.06 N = 15	14.84
Low	18.54 N = 11	14.58 N = 12	17.58 N = 12	18.13 N = 15	17.29
Total	16.34	14.65	16.78	16.10	16.01

A t-test analysis of interaction for the "high" group resulted in no differences between the means. The analysis for the "low" groups indicated that Form II (long without practice) was perceived as easier than the other three forms; i.e., than Form I, $t = 2.68$; $df = 21$; $p < .02$; than Form III, $t = 2.24$; $df = 22$; $p < .04$; than Form IV, $t = 2.43$; $df = 25$; $p < .03$.

DISCUSSION

An analysis of the results of achievement scores as a function of the number of associations per frame provided support for the findings in Experiment II that achievement was a function of the number of associations per frame, where consideration is given to the ability of the learner. Specifically, it was found that Ss of high ability did not differ significantly in their achievement as a function of the number of attributes per frame, while Ss of lower ability achieved significantly higher on the program that presented one attribute per frame. It is conceivable that similar results might have been found in Experiment I if the same precision in programming technique had been used.

The findings of this study were not entirely consistent with the findings of Experiment III. Although the above-average Ss achieved more on the easy program with practice than the long program without practice, significant differences for other similar comparisons for high and low ability as a function of practice were not obtained. These apparent inconsistencies with the findings of Experiment III may be explained by the fact that practice is meaningful only if acquisition has occurred. Thus, it could be concluded from this study that acquisition of the attributes of the concepts was low for (1) the low ability group on both long and short versions of the experimental programs, and (2) the high ability group on the short program.

An inspection of the data provides evidence that achievement was extremely low for both ability levels over all treatments, thus providing support for the hypothesis that if practice were to contribute to achievement, it would occur most likely for the high group on the easy program. It was observed that the high I.Q. group did achieve at a higher level on the easy program.

Additional support for this hypothesis can be found in Experiment III where achievement was high and practice had a significant effect on achievement.

An inspection of the mean number of responses made by Ss in a specified period of time indicates that practice for both ability levels reduces the average time required to make a response. A meaningful explanation for this may be that the direct recall of an attribute which was the probable method for responding to the practice frames, requires less time than does the cognitive process of associating attributes which was required for making a response to the learning frames. Thus, when an experimental program included practice frames, the mean number of responses which were made in a specified period of time was increased. An interesting observation was that the low ability Ss achieved significantly less than the high ability Ss on the achievement measures while they made a greater mean number of responses than the high ability group in the same period of time. From this, it could be concluded that the greater the amount of time the Ss spent in associating the attributes to be learned, the more effective the learning. The only data that would seem to be inconsistent with this conclusion would be the fact that both high and low ability groups made a smaller number of responses in the same period of time on their respective short forms of the program than they did on the long forms. It could be argued that if too many attributes are presented at one time, interference effects may occur, thus reducing the efficiency in responding.

Although the implications of these data are interesting, it should be pointed out that all conclusions based on data concerning the mean number of responses made in a specified period of time should be considered to be tentative, since the data were not treated statistically.

Since the F for efficiency scores as a function of differences in programs did not differ significantly, it could be concluded that efficiency is not affected by the number of associations per frame or practice of the attribute. The fact that efficiency scores differed as a function of ability probably reflects the significantly higher achievement scores of the high ability group.

The results of the analysis of the Ss perceived difficulty scores were generally consistent with findings in Experiment II. In most comparisons, the higher the Ss achievement on a program, the easier he perceived it. Specifically, the high ability Ss had significantly higher achievement scores than the low ability Ss and also perceived the programs as being easier than the low ability Ss. Further, the highest mean achievement score for the low ability groups was obtained on the long program without practice. This program was perceived by the Ss as significantly easier than any of the other three forms. The only exception to this was observed in the comparison of the high ability Ss. In this case, perceived difficulty scores did not differ significantly, while differences in achievement as a function of program differences did.

In conclusion, the findings of this study would indicate that achievement is a function of the number of associations per frame where consideration is given to the ability of the learner; that level of achievement is related to the practice of attributes during the learning of a concept if the initial acquisition level of the attributes is high; that achievement has some relationship to the amount of time the learner spends in making a response; and that the learner's perceived difficulty of a program is a function of his level of achievement on the program.

GENERAL CONCLUSIONS FOR THE FOUR EXPERIMENTS PRESENTED IN THIS REPORT

---

1. A system for defining frame difficulty based on stages of concept learning is effective.
2. Frame difficulty can be effectively manipulated in the association stage of concept learning.
3. The acquisition of the association stage of learning a concept is a function of the number of associations of attributes per frame, the number of associations of attributes per concept, and the level of student ability.
4. Achievement of students with above-average ability is reduced if the number of associations of attributes per frame is either too small or too great.
5. Achievement of students with below-average ability is maximized if the number of attributes presented per frame is limited in number.
6. Practice of an attribute after its acquisition increases the effectiveness of acquiring the association stage of concept learning for students with either above- or below-average ability.
7. For practice of an attribute to contribute to the acquisition of the association stage of concept learning acquisition of the attribute must have first occurred.
8. Students' perception of concept difficulty is a function of the interaction of the number of associations of attributes per frame and the level of student ability.
9. Students' perception of concept difficulty is a function of the interaction of the number of associations of attributes per concept and the level of student ability.

BIBLIOGRAPHY

Angell, D., and A. A. Lumsdaine. The Effects of Prompting Trials and Partial Correction Procedures on Learning by Anticipation. American Institute for Research, San Mateo, Calif. AIR-C14-9/61-SR5.

Beane, D.G. "A Comparison of Linear and Branching Techniques of Programmed Instruction in Plane Geometry". Urbana, Ill., Univ. of Illinois, Trng. Res. Lab., USOE Title VII, Project 711151.01, Tech. Rep. No. 1, (Doctoral Dissertation), July, 1962.

Campbell, V. N. Adjusting self-instruction programs to individual differences: studies of cueing, responding and by-passing. American Institute for Research, San Mateo, Calif. AIR-C41-7/61-SR. p. 13.

Carr, W. J. "A Functional Analysis of Self-Instructional Devices". In Teaching Machines and Programmed Learning: A Source Book, A. A. Lumsdaine and R. Glaser (eds.) NEA, Washington, D. C. pp. 540-562.

Cline, O. B., A.R. Beals, and D. Seidman. Experimenting with accelerated training programs for men of various intelligence levels. Educ. Psychol. Measurement, 20:723-735, 1960.

Coulson, J.E., D. P. Estavan, R. J. Melagarno, and H.F. Silberman. Effects of branching in a computer controlled auto-instructional device. J. Applied Psychol., 46(6):389-392, 1962.

Coulson, J.E., and H.F. Silberman. Results of an initial experiment in automated teaching. J. Educ. Psychol., 51:135-143, 1960.

DeCecco, J.P. Educational Technology. New York: Holt, Rinehart and Winston, Inc., p. 346, 1964.

Evans, J.L., R. Glaser, and L.E. Homme. An investigation of variation in the properties of verbal learning sequences of the "teaching machine" type. In A.A. Lumsdaine, and R. Glaser, Teaching Machines and Programmed Learning: A Source Book. NEA, Washington, D.C. 1960. pp. 446-451.

Gagné, R.M. The Conditions of Learning. New York: Holt, Rinehart and Winston, Inc., 1965. Chapters 4 and 5.

Glaser, R., and J. H. Reynolds. "Multi-tracking in programmed instruction: studies with a program on mathematical bases for management decision-making". Pittsburgh, Pa.: Programmed Learning Laboratory, Univ. of Pittsburgh. (1962).

Green, E.J. The Learning Process and Programmed Instruction. New York: Holt, Rinehart and Winston, Inc., 1962. p. 188.

Homme, L.E., R. Glaser. Relationships between the programmed textbook and teaching machines. In Automatic Teaching: The State of the Art., E.H. Galanter (ed.) New York: John Wiley and Sons, Inc., 1959. Chap. 9, pp. 103-108.

Klausmeier, H., and P. Lambert. Teaching Machines and the Learning Process. Educ. Leadership, 1961, 13(4):278-283.

Lumsdaine, A.A. Some Issues Concerning Devices and Programs for Automated Learning. In Teaching Machines and Programmed Learning: A Source Book, A.A. Lumsdaine and R. Glaser (eds.) NEA, Washington, D.C. pp. 517-539.

Lumsdaine, A.A. "Some Theoretical and Practical Problems in Programmed Instruction". In DeCecco, J.P., Educational Technology - Readings in Programmed Instruction. New York: Holt, Rinehart and Winston, Inc. 1964. p. 82.

Maccoby, N., and F. D. Sheffield. Theory and experimental research on the teaching of complex sequential procedures by alternate demonstration and practice. In Finch, G., and F. Cameron (eds.) Symposium on Air Force Human Engineering, Personnel and Training Research (Washington, D.C.: Nat. Acad. of Sci., Nat. Res. Council, 1958).

Margolius, G.J., and F. D. Sheffield. Optimum methods of combining practice teaching with filmed demonstration teaching complex response sequences: serial learning of a mechanical assembly task. In A.A. Lumsdaine (ed.), Student Response in Programmed Instruction: A symposium on experimental studies of cue and response factors in group and individual learning from instructional media. Washington, D.C.: Nat. Acad. Sci., Nat. Res. Council (Pub. 943); sponsored by Hdqtrs., ARDC, USAF. 1961.

Moore, J.W., W.I. Smith and R.C. Teevan. Motivational Variables in Programmed Learning. Monograph. Final Report, USOE, Title VII, Grant No. 7-48-0070-149.1. 1965.

Porter, D. "Teaching Machines". In Teaching Machines and Programmed Learning: A Source Book, A.A. Lumsdaine and R. Glaser (eds.) NEA, Washington, D.C. pp. 209-210.

Roe, A. A comparison of branching methods for programmed learning. J. Educ. Res. 55:407-416. 1962.

Shay, C.B. Relationship of intelligence to step size on a teaching machine program. J. Educ. Psychol. 52(2):98-104. 1961.

Silberman, H.F., R. J. Melagarno, J.E. Coulson, and D. Estavan. Fixed sequence vs. branching auto-instructional methods. J. Educ. Psychol., 52:166-172, 1961.

Skinner, B.F. Teaching Machines. Science, 128:969-977, 1958.

Smith, W.I. and J.W. Moore. Size-of-step and cueing. Psychol. Rep. 10:287-294. 1962.

Trow, W.C. Teacher and Technology, New Designs for Learning. New York: Meredith Publishing Co. 102 pp. 1963.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Co. 1962.